ZooKeys 1129: 55–72 (2022)
doi: 10.3897/zookeys.1129.91543
https://zookeys.pensoft.net

RESEARCH ARTICLE

A peer-reviewed open-access journal
ZooKeys
Launched to accelerate biodiversity research

Genome-wide survey reveals the phylogenomic relationships of *Chirolophis japonicus* Herzenstein, 1890 (Stichaeidae, Perciformes)

Lu Liu[1], Qi Liu[2], Tianxiang Gao[3,4]

1 *Naval Architecture and Port Engineering College, Shandong Jiaotong University, Weihai, China* **2** *Wuhan Onemore-tech Co., Ltd. Wuhan, Hubei, China* **3** *Fishery College, Zhejiang Ocean University, Zhoushan, Zhejiang, China* **4** *Zhejiang Provincial Key Laboratory of Mariculture and Enhancement, Zhejiang Marine Fisheries Research Institute, Zhoushan, China*

Corresponding author: Tianxiang Gao (gaotianxiang0611@163.com)

Academic editor: Maria Elina Bichuette | Received 11 August 2022 | Accepted 6 October 2022 | Published 11 November 2022

https://zoobank.org/B99966E1-EF69-46FD-8CDF-D6D1417102C1

Citation: Liu L, Liu Q, Gao T (2022) Genome-wide survey reveals the phylogenomic relationships of *Chirolophis japonicus* Herzenstein, 1890 (Stichaeidae, Perciformes). ZooKeys 1129: 55–72. https://doi.org/10.3897/zookeys.1129.91543

Abstract

Fish are the largest vertebrate group, consisting of more than 30 000 species with important ecological and economical value, while less than 3% of fish genomes have been published. Herein, a fish, *Chirolophis japonicus*, was sequenced using the next-generation sequencing. Approximately 595.7 megabase pair of the *C. japonicus* genome was assembled (49 901 contigs with 42.61% GC contents), leading to a prediction of 46 729 protein-coding gene models. A total of 554 136 simple sequence repeats was identified in the whole genome of *C. japonicus*, and dinucleotide microsatellite motifs were the most abundant, accounting for 59.49%. Phylogenomic analysis of 16 genomes based on the 694 single-copy genes suggests that *C. japonicus* is closely related with *Anarrhichthys ocellatus*, *Cebidichthys violaceus*, and *Pholis gunnellus*. The results provide more thorough genetic information of *C. japonicus* and a theoretical basis and reference for further genome-wide analysis.

Keywords

Chirolophinae, draft genome, genome assembly, genome evolution, next-generation sequencing, Stichaeidae, Zoarcales

Copyright *Lu Liu et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

Chirolophis Swainson, 1839 belongs to the family Stichaeidae of the order Perciformes, which is widely distributed between cold and temperate areas in the Pacific Ocean and along the coasts of Europe in the Atlantic Ocean (Jing et al. 2005; Balanov et al. 2020). *Chirolophis* contains nine species (https://www.fishbase.se/search.php) which are important commercial bony fishes, especially in China (Chen et al. 2017). Among these species, *Chirolophis japonicus* (Herzenstein, 1890), also known as *Azuma emmnion* (Jordan & Snyder, 1902), lives in rocky shallow coastal waters of the Pacific Ocean, including the Yellow Sea, the Bohai Sea, the northern Sea of Japan, and the Okhotsk Sea to the Bering Sea (Shiogaki 1983; Jing et al. 2005; Balanov et al. 2020). They display strong cryptic habits and are almost impossible to be observed by SCUBA diving observations. Studies on this species are relatively rare, mainly including mitochondrial genome data (Yang et al. 2016), the origin of the cortical protrusion of head (Sato 1977), and reproductive biology research (Chen 2017).

Genome-based phylogenetic studies have provided new opportunities for exploring the phylogeny of fishes. With the development of molecular biology and sequencing technology, more and more species are being sequenced and genomes published, ranging from model fishes to many commercial species. There are nearly 9900 species published genomes in the Eukaryota on the NCBI database (https://www.ncbi.nlm.nih.gov/genome/), accessed on 7 July 2022. Genome survey sequencing (GSS) was considered useful for providing basic genome information. Besides productively identifying genome-wide simple sequence repeats (SSRs) effectively, it can predict putative gene functions efficiently and target the potential exon-intron boundaries. A series of research advances has been made in the study of phylogenomic relationships of organisms, such as plants (Ran et al. 2018; Li et al. 2019b), animals (Koepfli et al. 2015; Heras et al. 2020), and fungi (Spatafora et al. 2016; Liu et al. 2022), which have provided insight into evolutionary history.

In the order Perciformes, the genomes of only three species, *Anarrhichthys ocellatus* (Ayres, 1855), *Cebidichthys violaceus* (Girard, 1854), and *Pholis gunnellus* (Linnaeus, 1758), have been published so far (Li et al. 2019a; Heras et al. 2020; Potter and Consortium 2022). Meanwhile, the complete mitogenomes of two species, *Chirolophis ascanii* and *Chirolophis japonicus* (or *Azuma emmnion*), provided robust phylogenetic relationships (Yang et al. 2016; Chen et al. 2017; Margaryan et al. 2021). Completed genome sequences of *C. japonicus* would improve our understanding of phylogeny, even though the genomic information of *C. japonicus* remains unknown.

In this present study, we perform a genomic survey for *C. japonicus* using next-generation sequencing technology for the first time, investigate its genomic feature and reconstruct the phylogenomic relationships with single-copy orthologs genes of *C. japonicus*. The draft genome assembly of *C. japonicus* can help us find more useful information for taxonomic studies, adaptive evolutionary mechanisms, and phylogenetic studies, as well as understand the genomic evolution of *Chirolophis*, and provide a molecular basis of *C. japonicus*.

Materials and methods

Material collection

In this study, a male specimen of *C. japonicus* with body length 186 mm and body weight 225 g was collected from coastal waters of Qingdao (35°40'N, 119°30'E), China in July 2021 (Fig. 1). Firstly, we identified it by morphological characteristics and DNA barcoding (mitochondrial DNA COI gene), then the examined sample was quickly preserved in −80 °C ultra-low temperature freezer. All subsequent animal experiments took place at Fisheries Ecology and Biodiversity Laboratory (**FEBL**) of Zhejiang Ocean University, Zhoushan, China. Experiments were conducted under the guidelines and approval of the Ethics Committee for Animal Experimentation of Zhejiang Ocean University (ZJOU-ECAE20211876). Secondly, a piece of fresh muscle tissue was clipped from the base of dorsal fin and preserved in absolute 95% ethanol.

Genomic DNA extraction and next-generation sequencing

The total cell DNA was extracted using the phenol-chloroform method (Sambrook et al. 1982), following the protocol in a previous study (Yang et al. 2021), and then carried out with DNA/Protein Analyzer and 1% agarose gel electrophoresis. High-quality DNA was randomly interrupted using ultrasonic crusher, and the obtained short reads (300–350 bp) were sequenced with Illumina NovaSeq 6000 with a paired-end library following the manufacturer's instructions (OneMore-Tech, Wuhan, China) in January 2022.

Sequence quality control, genome assembly, and K-mer analysis

Quality control was performed on the raw data from the Illumina sequencing platform using the FastQC v. 0.11.9 (Andrews 2010) and Trimmomatic v. 0.39 (Bolger et al. 2014) based on four criteria: 1) removal of the A-tail and adaptors, 2) deletion of the low-quality reads where N contents are more than 10%, 3) filtration of the reads whose base quality is less than 10, and 4) discard of duplicated reads. The genome size, heterozygosity, and repeat content of *C. japonicus* was estimated based on a K-mer method (Liu et al. 2013). De novo assembly of the *C. japonicus* genome was conducted using MaSuRCA v. 3.3.3 (Zimin et al. 2013) based on clean data.

Figure 1. *Chirolophis japonicus* (Herzenstein, 1890), 186 mm, from Qingdao.

The quality of the assembled genome was evaluated by Quast v. 5.0.2 and BUSCO v. 5.3.2 (Simão et al. 2015). The mitochondrial DNA analyses followed the method of previous studies (Yang et al. 2016, 2021; Nie et al. 2021). In brief, the software NOVOPlasty v. 4.2.1 (Dierckxsens et al. 2017) and GetOrganelle v. 1.7.6.1 (Jin et al. 2020) were used to assemble the mitogenome with clean data. The mitogenome of *C. japonicus* was annotated using MFannot tool (http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl) and GeSeq (Tillich et al. 2017), then manually annotated and drawn with OGDraw v. 1.3.1 (Lohse et al. 2013; Greiner et al. 2019). The clean data and complete assembled mitochondrial genome were uploaded to GenBank.

Gene prediction and functional annotation

The gene predictors Augustus v. 3.3.3 (Stanke and Waack 2003), SNAP (Johnson et al. 2008), and GeneMark-ES v. 4.69 (Lomsadze et al. 2005) were trained on the gene models, and all the gene models were integrated using EvidenceModeler v. 1.1.1 (Haas et al. 2008). The amino acid sequences from *C. japonicus* were annotated by GO (Ashburner et al. 2000), Eggnog (Huerta-Cepas et al. 2019), CAZymes (Cantarel et al. 2009), InterPro (Hunter et al. 2009), KEGG (Kanehisa et al. 2006), KOG https://www.creative-proteomics.com/services/kog-annotation-analysis-service.html, and Pfam (El-Gebali et al. 2019), using Diamond v. 2.0.2 with the e-value less than 1×10^{-5} (Buchfink et al. 2015).

Microsatellite identification and non-coding RNA annotation

In this study, MIcroSAtellite identification tool (MISA) v. 2.1 was used to identify simple sequence repeats (SSR) in the draft genome of *C. japonicus* (Thiel et al. 2003). The tRNA and rRNA were predicted by tRNAscan-SE v. 3.0 (Lowe and Eddy 1997) and RNAmmer v. 1.2 (Lagesen et al. 2007), respectively.

Phylogenomic analysis of *C. japonicus*

A total of 15 genomes of other bony fish were downloaded from the NCBI database (Table 1). The amino acid sequences of single-copy orthologs genes among the 16 species were found using OrthoFinder v. 2.5.4 (Emms and Kelly 2019), and these sequences were aligned by using MAFFT v. 7 (Katoh and Standley 2013). In order to reconstruct the phylogenomic relationship of *C. japonicus*, a maximum likelihood (ML) tree was analyzed/constructed using RaxML v. 8.2.12 based on the amino acid sequences of single-copy orthologs genes (Stamatakis 2014). The best model was PROTGAMMAILGF with 100 bootstrap replicates. Finally, the phylogram was viewed using FigTree v. 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

Table 1. Information on genomes used in this study.

Species	Biosample	Bioproject	References
Anarrhichthys ocellatus	SAMN10245424	PRJNA496475	
Archocentrus centrarchus	SAMN09948522	PRJNA489129	Koepfli et al. 2015
Cebidichthys violaceus	SAMN06857690	PRJNA384078	Heras et al. 2020
Chirolophis japonicus			**This study**
Cyclopterus lumpus	SAMN12629502	PRJNA625538	
Gasterosteus aculeatus	SAMN15223905	PRJNA707557	Berner et al. 2019; Nath et al. 2021
Gymnodraco acuticeps	SAMEA104242997	PRJEB37639	
Liparis tanakae	SAMN10970109	PRJNA523297	
Micropterus salmoides	SAMN15299117	PRJNA687018	Broughton and Reneau 2006; Sun et al. 2021
Myoxocephalus scorpius	SAMEA4028818	PRJEB12469	
Pholis gunnellus	SAMEA7522838	PRJEB45449	
Pseudoliparis sp.	SAMN10662039	PRJNA512070	Mu et al. 2021
Seriola lalandi	SAMN04902367	PRJNA319656	Purcell et al. 2018
Taurulus bubalis	SAMEA7522994	PRJEB45317	
Toxotes jaculatrix	SAMN18445299	PRJNA723051	
Ophiodon elongatus	SAMN13559843	PRJNA595583	Longo et al. 2020

Data availability statement

Raw sequencing data for genome have been deposited at the Sequence Read Archive SRR21530970. These data can be quickly accessed by checking the project ID PRJNA879413 at NCBI Project.

Results

Sequencing data statistics and K-mer analysis

In this study, a total of 65.4 Gb clean reads was obtained by next-generation sequencing from an Illumina NovaSeq 6000 platform. The Q20 value, Q30 value, and GC content were 98.17%, 94.83%, and 43.14%, respectively. The K-mer analysis with a depth of 71 shows that genome size of *C. japonicus* was 596 Mb with 0.50% heterozygosity rate and 30.30% repeat sequences (Table 2, Suppl. material 1), resulting in *C. japonicus* being a diploid.

Genomic and mitochondrial features

The genome sequences of *C. japonicus* were sequenced from a male with an Illumina NovaSeq 6000 platform, spanning 595.7 Mb with GC contents of 42.61% that were assembled using the software MaSuRCA (Table 3; Zimin et al. 2013). A total of 49

Table 2. The genome characteristics of *Chirolophis japonicus* based on the K-mer method.

Species	K-mer number	K-mer depth	Genome size (Mb)	Heterozygous ratio (%)	Repeat sequences (%)
C. japonicus	4.353×10^{10}	71	596	0.50	30.30

Table 3. Gene prediction and annotation of *Chirolophis japonicus*.

Category	Database	Number of reads	Percent (%)
Protein-coding gene model		46 729	
Annotated	InterPro	37 169	79.54
	Eggnog	37 742	80.98
	GO	9353	20.02
	KEGG_KO	17 747	37.98
	Pfam	26 530	56.77
	KOG	35 440	75.84
	CAZymes	765	1.64
Assembly BUSCO coverage			88.9

901 contigs was generated with the largest contigs of 365 029 bp. The final contigs N50 and L50 were 29 108 bp and 5388 bp long, respectively (Table 3). A total of 69 rRNA was identified, including 66 8S rRNA, two 18S rRNA, and single 28S rRNA. In addition, 846 tRNA were annotated using the tRNAscan-SE.

The complete mitogenome of *C. japonicus* is 16,522 bp long with a GC content of 45.97%. It consists of two ribosomal RNA genes (rnl and rns), 20 tRNA genes, and 13 protein-coding genes (PCGs) without an intron (Fig. 2).

Chirolophis japonicus genome annotation

A total of 46 729 protein-coding genes was predicted by a combination of different software, including Augustus v. 3.3.3 (Stanke and Waack 2003), SNAP (Johnson et al. 2008) and GeneMark-ES v4.69 (Lomsadze et al. 2005). Among these, 79.54%, 80.98%, 20.02%, 39.98%, 56.77%, 75.84%, and 1.64% genes were annotated in the InterPro, Eggnog, GO, KEGG_KO, Pfam, KOG, and CAZymes databases, respectively.

Distribution and features of SSR

A total of 554 136 of SSR was identified in the complete genome of *C. japonicus*, including 166 077 of mononucleotide microsatellite motifs (29.97%), 329 685 of dinucleotide microsatellite motifs (59.49%), 37 615 of trinucleotide microsatellite motifs (6.79%), 17 896 of tetranucleotide microsatellite motifs (3.23%), 1568 of pentanucleotide microsatellite motifs (0.28%), and 1322 of hexanucleotide microsatellite motifs (0.24%;) (Fig. 2). A/T, AC, GAG, AGAC, CTCTC, and CCCTAA were the highest repeats in mono-, di-, tri-, tetra-, penta-, and hexanucleotide microsatellite motifs, respectively (Fig. 3).

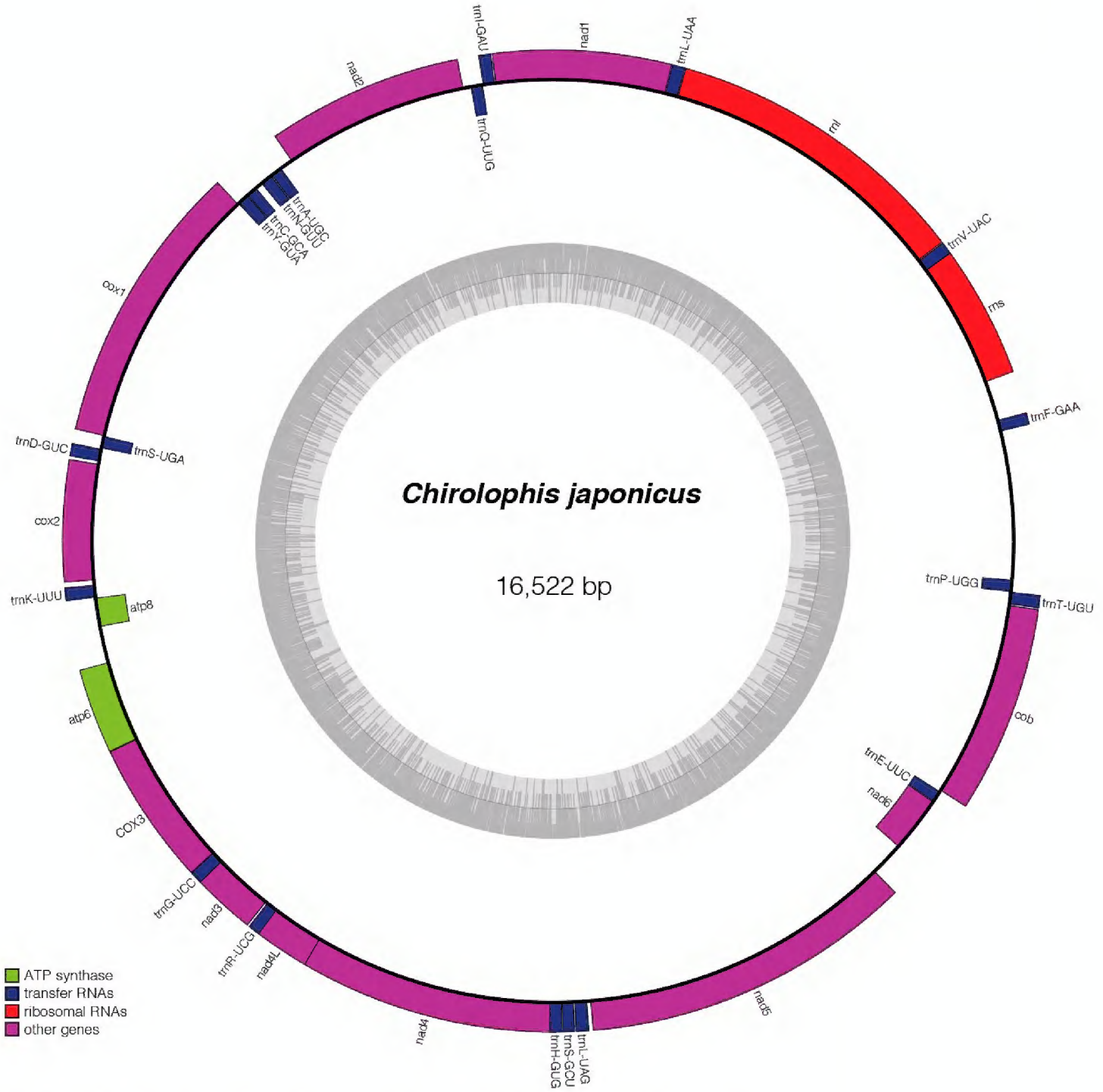

Figure 2. The complete mitogenome structure of *Chirolophis japonicus*.

Phylogenomic relationships of *Chirolophis japonicus*

In the present study, the phylogenomic relationship of a total of 16 bony fish (Table 1, Fig. 4) was reconstructed. A total of 694 single copy genes (Suppl. material 2) was identified from 16 male fish genomes using OrthoFinder (Emms and Kelly 2019), which consisted of 361 031 characters from amino acid sequences. The phylogenomic tree suggested that *Chirolophis japonicus* is closely related with *Anarrhichthys ocellatus*, *Cebidichthys violaceus*, and *Pholis gunnellus*, and provided robust phylogenetic relationships within the order Zoarcales, with full support. Although *Chirolophis japonicus* and *Cebidichthys violaceus* belong to the family Stichaeidae, they did not form a clade based on the amino acid sequences of 694 single-copy genes. In addition, *Ophiodon elongatus*, *Cyclopterus lumpus*, *Liparis tanakae*, *Pseudoliparis* sp., *Taurulus bubalis*, and *Myoxocephalus scorpius* clustered into a clade; other species, including *Gymnodraco acuticeps*, *Archocentrus*

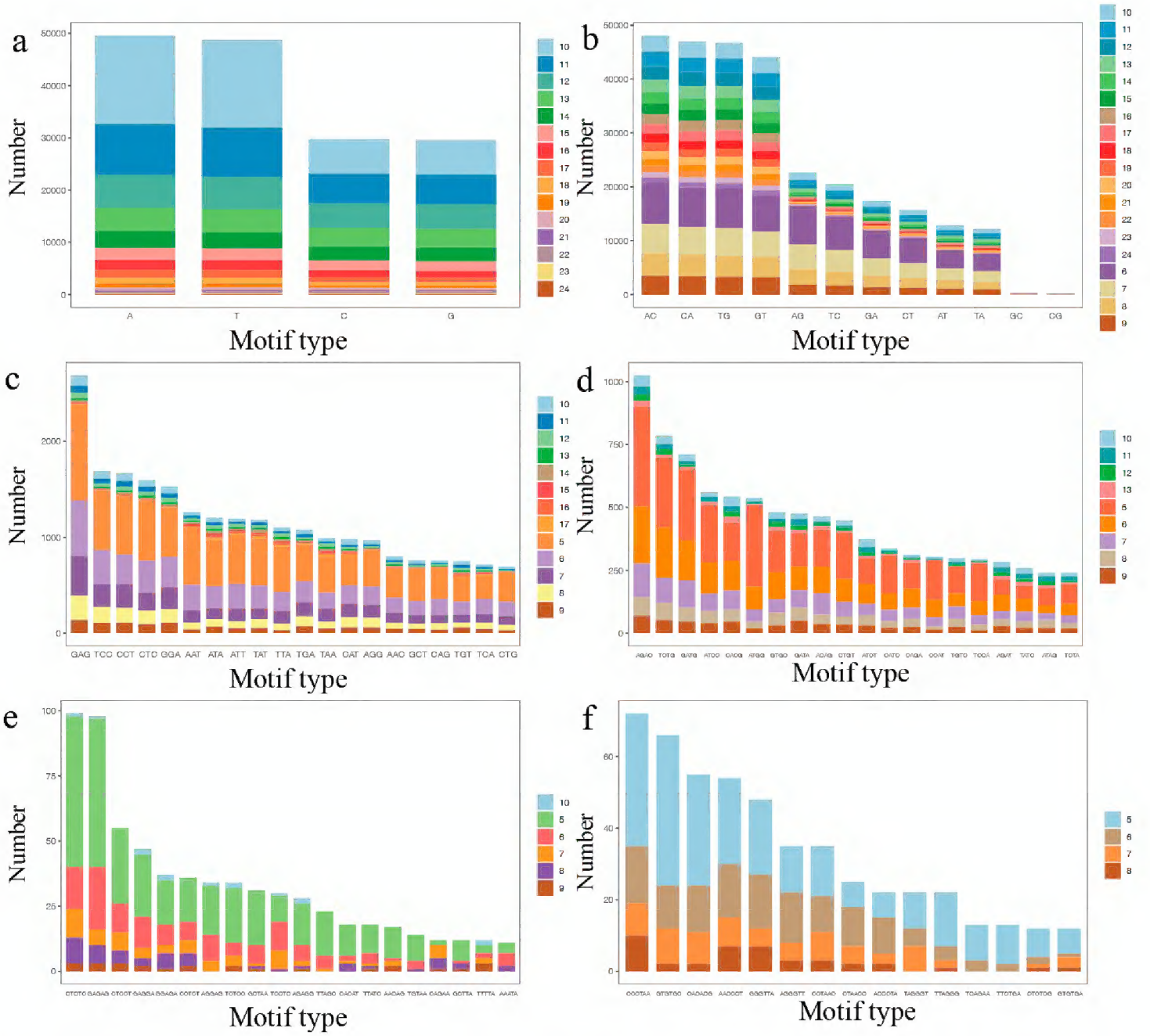

Figure 3. The distributions and frequencies of microsatellite motifs of *Chirolophis japonicus* **a** mononucleotide microsatellite motifs **b** dinucleotide microsatellite motifs **c** trinucleotide microsatellite motifs **d** tetranucleotide microsatellite motifs **e** pentanucleotide microsatellite motifs **f** hexanucleotide microsatellite motifs.

centrarchus, *Seriola lalandi*, *Toxotes jaculatrix*, *Micropterus salmoides* and *Gasterosteus aculeatus*, formed a separate clade. In addition, the phylogenomic analysis based on the amnio acid of 13 protein-coding genes of mitogenome show that *Chirolophis japonicus* is closely related with the *Chirolophis ascanii* (Fig. 5).

Discussion

Currently, there are more than 30 000 species of fishes, including bony, jawless, and cartilaginous fishes, living on the earth, some with great ecological and economic value. In 2002, the first fish genome, *Fugu rubripes* (also known as "torafugu") was published, which provided a framework for future studies of fish genomes (Aparicio

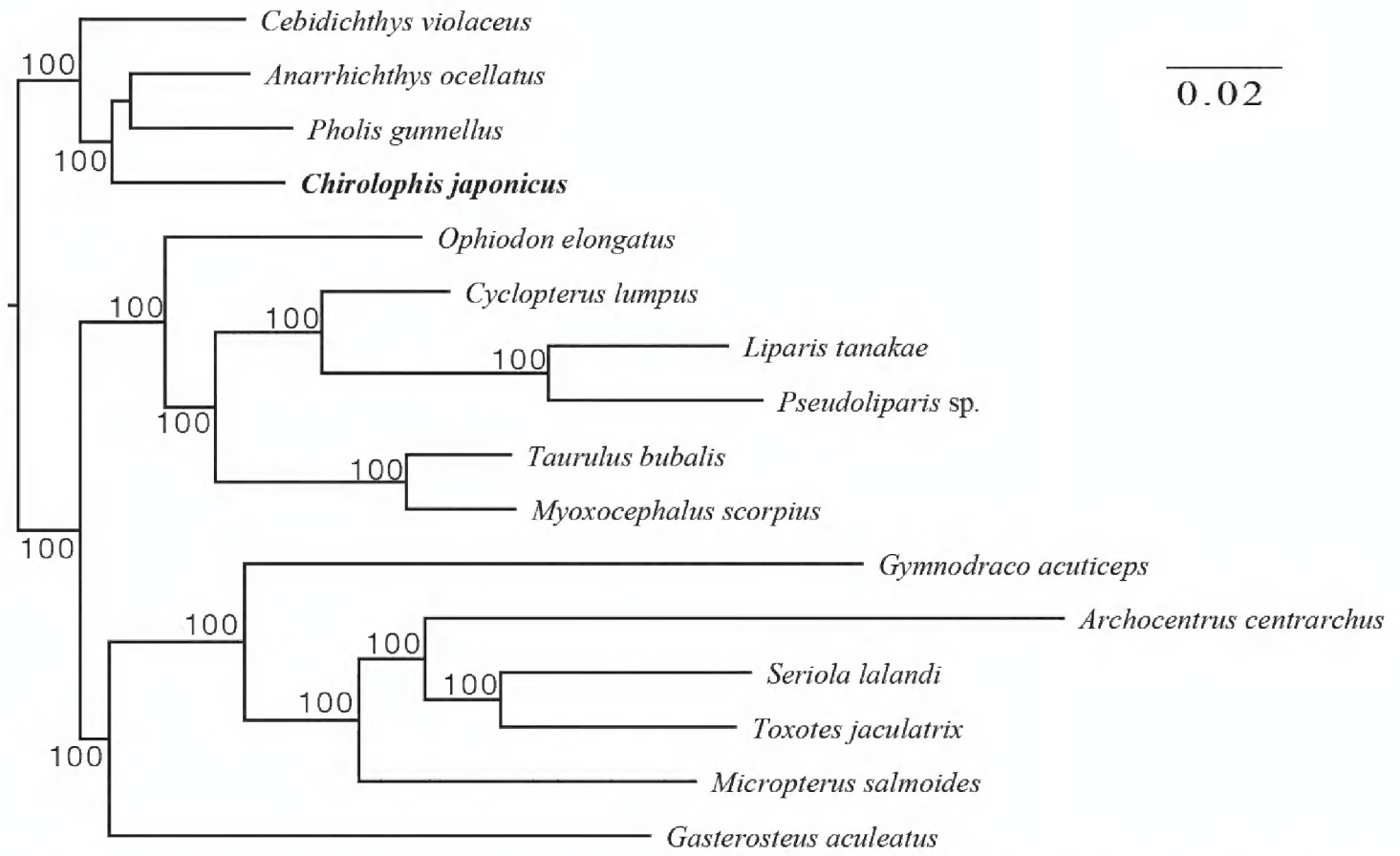

Figure 4. A maximum likelihood (ML) phylogenomic tree of *Chirolophis japonicus* based on amino acid sequences of 694 single-copy genes. *Chirolophis japonicus* is in bold. Maximum likelihood bootstrap values (90%) of each clade are indicated along branches. A scale bar in the upper right indicates substitutions per site.

et al. 2002). With the rapid development of the whole genome sequencing (WGS) technology, a large number of fish genomes have since been sequenced, such as fish as model organisms *Oryzias latipes* and *Danio rerio*, and economically important fishes such as *Cyprinus carpio* and *Ctenopharyngodon idella* (Kasahara et al. 2007; Howe et al. 2013; Xu et al. 2014; Wu et al. 2022). In addition, the Chinese "Aquatic 10-100-1000 Ge- nomics Program" and the "Fish 10K Project" have facilitated the understanding of fish genomes (Liu et al. 2017; Fan et al. 2020). Until now, a total of 819 fish genomes has been released in the NCBI database (https://www.ncbi.nlm.nih.gov/genome/, assessed on 7 July 2022), which is less than3% of the known 30 000 species.

In the present study, a new fish genome, *Chirolophis japonicus*, was sequenced. The genomes size was estimated to be 596 Mb based on the K-mer analysis, and the genome spanned 595.7 Mb, assembled using the MaSuRCA (Table 3; Zimin et al. 2013), which followed the predicted genome size of the K-mer method. Among the published teleost genomes, the size ranges from 322.5 Mb (*Fugu rubripes*) to 40 Gb (*Protopterus annectens*) (Aparicio et al. 2002; Wang et al. 2021), with an average length less than 1 Gb (Fan et al. 2020). Meanwhile, the genome of three species in Zoarcales, including *Anarrhichthys ocellatus* (612.19 Mb), three genomes of *Cebidichthys violaceus* (575.66 Mb, 593.00 Mb, 606.18 Mb), and two genomes of *Pholis gunnellus* (588.7 Mb, 590.3 Mb), are slightly larger than that of *C. japonicus* (https://www.ncbi.nlm.nih.gov/genome/, assessed on 7 July 2022). In addition, the heterozygous ratio of

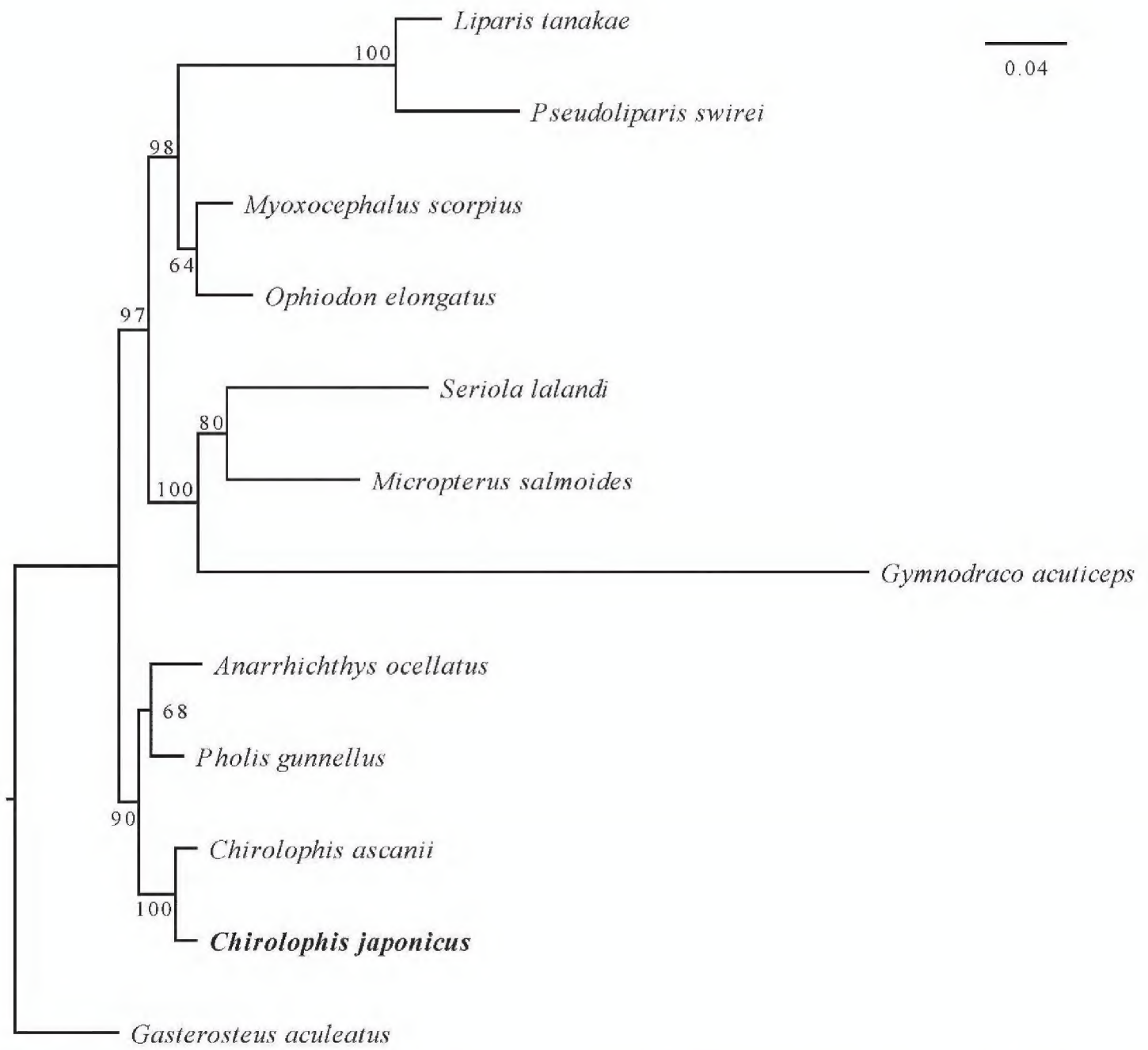

Figure 5. The maximum likelihood (ML) phylogenomic tree of fungi based on amino acid of 13 protein-coding genes (PCGs): ATP6, ATP8, COX1, COX2, COX3, CYTB, ND1, ND2, ND3, ND4, ND4L, ND5 and ND6. Support values for ML analysis greater than 60% is given on relative clade. A scale bar in the upper left indicates substitutions per site.

C. japonicus was 0.50%, probably mid-level compared to other teleost genomes (Aird et al. 2011; Li et al.2019c; Xu et al. 2020; Yang et al. 2021).

At present, phylogenomic analysis has become an important method for studying the evolutionary relationships of an organism, such as plants (Ran et al. 2018; Li et al. 2019b), animals (Koepfli et al. 2015; Heras et al. 2020), and fungi (Spatafora et al. 2016; Liu et al. 2022). Although the phylogenetic relationships of the genus *Chirolophis* have been published based on the mitogenomes (Yang et al. 2016; Chen et al. 2017; Margaryan et al. 2021), we provided a phylogenomic relationship according to the 694 single-copy genes (Fig. 3, Suppl. material 2) among *C. japonicus* and 15 other species. The results of the phylogenomic tree shows that *C. japonicus* is closely related with three species in the order Zoarcales, while *C. japonicus* and *Cebidichthys violaceus*, belonging to the family Stichaeidae,

are without a clade (Fig. 2). Thus, solving this problem requires more fish genomes to be sequenced.

Microsatellite DNA markers shows many advantages, such as codominant, extensive distribution, abundant polymorphisms, and a convenient analysis, and was considered to be an effective tool in genetic analysis and evolutionary research (Yang et al. 2022). In this study, the highest number and type of repeats is dinucleotide repeats, which was consistent with data for *Ophichthus evermanni* (Yang et al. 2022), *Padon nehereus* (Yang et al. 2021), *Cociella crocodilus* (Zhao et al. 2021), *Acanthogobius ommaturus* (Chen et al. 2020), *Sillago sihama* (Qiu et al. 2020), and other species. SSR polymorphic loci are mainly distributed among mononucleotide and dinucleotide repeats. Based on this, the search of polymorphic SSR markers from low repetitive motifs will greatly help in subsequent population genetics research of *C. japonicus*. The complexity of repeated motif usually reflects DNA mutation rate and evolutionary level (Katti et al. 2001). The frequency from mononucleotides to trinucleotides was up to 96.25%, which implies that *C. japonicus* has experienced a long evolutionary history and accumulated more genetic variation.

Finally, the genome assembly of *C. japonicus* can help us understand the genome evolution of *Chirolophis* and teleosts, as well as provide a molecular basis for breeding and cultivation.

Acknowledgements

We sincerely thank the reviewers for their constructive comments. We would like to thank Yuan Zhang and Chenghao Jia for assistance in sample collection and sorting. This work was supported by the Zhejiang Provincial Key Research and Development Program (2021C02047); The Doctoral Research Foundation of Shandong Jiaotong University, Grant/Award Numbers: BS201902051.

References

Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biology 12(2): 1–14. https://doi.org/10.1186/gb-2011-12-2-r18

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Date last accessed: April 18, 2020]

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MDS, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJK, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297(5585): 1301–1310. https://doi.org/10.1126/science.1072104

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: Tool for the unification of biology. Nature Genetics 25(1): 25–29. https://doi.org/10.1038/75556

Balanov A, Epur I, Shelekhov V (2020) A Description of *Chirolophis japonicus* and *Ch. saitone* (Stichaeidae) Pelagic Larvae from Peter the Great Bay, Sea of Japan. Journal of Ichthyology 60(3): 364–374. https://doi.org/10.1134/S0032945220030042

Berner D, Roesti M, Bilobram S, Chan SK, Kirk H, Pandoh P, Taylor GA, Zhao Y, Jones SJM, DeFaveri J (2019) De novo sequencing, assembly, and annotation of four three spine stickleback genomes based on microfluidic partitioned DNA libraries. Genes 10(6): 426. https://doi.org/10.3390/genes10060426

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30(15): 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Broughton RE, Reneau PC (2006) Spatial covariation of mutation and nonsynonymous substitution rates in vertebrate mitochondrial genomes. Molecular Biology and Evolution 23(8): 1516–1524. https://doi.org/10.1093/molbev/msl013

Buchfink B, Xie C, Huson DH (2015) Fast and sensitive protein alignment using DIAMOND. Nature Methods 12(1): 59–60. https://doi.org/10.1038/nmeth.3176

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Research 37(Database): 233–238. https://doi.org/10.1093/nar/gkn663

Chen F (2017) Study on the biological characteristics of *Azuma emmnion*. Master thesis, Dalian Ocean University, Dalian, China.

Chen X, Chen Y, Yu M, Sha Z, Shan X (2017) The complete mitochondrial genome of the *Azuma emmnion*. Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis 28(1): 77–78. https://doi.org/10.3109/19401736.2015.1110806

Chen B, Sun Z, Lou F, Gao T, Song N (2020) Genomic characteristics and profile of microsatellite primers for *Acanthogobius ommaturus* by genome survey sequencing. Bioscience Reports 40(11): 1–8. https://doi.org/10.1042/BSR20201295

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Research 45: e18. https://doi.org/10.1093/nar/gkw955

El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD (2019) The Pfam protein families database in 2019. Nucleic Acids Research 47(D1): D427–D432. https://doi.org/10.1093/nar/gky995

Emms DM, Kelly S (2019) OrthoFinder: Phylogenetic orthology inference for comparative genomics. Genome Biology 20(1): 238. https://doi.org/10.1186/s13059-019-1832-y

Fan G, Song Y, Yang L, Huang X, Zhang S, Zhang M, Yang X, Chang Y, Zhang H, Li Y, Liu S, Yu L, Chu J, Seim I, Feng C, Near TJ, Wing RA, Wang W, Wang K, Wang J, Xu X, Yang H, Liu X, Chen N, He S (2020) Initial data release and announcement of the 10,000 Fish Genomes Project (Fish10K). GigaScience 9(8): giaa0080. https://doi.org/10.1093/gigascience/giaa080

Greiner S, Lehwark P, Bock R (2019) Organellar Genome DRAW (OGDRAW) version 1.3. 1: Expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Research 47(W1): W59–W64. https://doi.org/10.1093/nar/gkz238

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR (2008) Automated eukaryotic gene structure annotation using evidence modeler and the program to assemble spliced alignments. Genome Biology 9(1): 1–22. https://doi.org/10.1186/gb-2008-9-1-r7

Heras J, Chakraborty M, Emerson JJ, German DP (2020) Genomic and biochemical evidence of dietary adaptation in a marine herbivorous fish. Proceedings. Biological Sciences 287(1921): e20192327. https://doi.org/10.1098/rspb.2019.2327

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch G-J, White S, Chow W, Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, Vogel J-H, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Elliott D, Threadgold G, Harden G, Ware D, Begum S, Mortimore B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Lloyd C, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper JD, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Ürün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Lanz C, Raddatz G, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Schuster SC, Carter NP, Harrow J, Ning Z, Herrero J, Searle SMJ, Enright A, Geisler R, Plasterk RHA, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nüsslein-Volhard C, Hubbard TJP, Crollius HR, Rogers J, Stemple DL (2013) The zebrafish reference genome sequence and its relationship to the human genome. Nature 496(7446): 498–503. https://doi.org/10.1038/nature12111

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P (2019) eggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research 47(D1): D309–D314. https://doi.org/10.1093/nar/gky1085

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L, Finn RD, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Laugraud A, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, Mistry J,

Mitchell A, Mulder N, Natale D, Orengo C, Quinn AF, Selengut JD, Sigrist CJA, Thimma M, Thomas PD, Valentin F, Wilson D, Wu CH, Yeats C (2009) InterPro: The integrative protein signature database. Nucleic Acids Research 37(Database): D211–D215. https://doi.org/10.1093/nar/gkn785

Jin JJ, Yu WB, Yang JB, Song Y, DePamphilis CW, Yi TS, Li DZ (2020) GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biology 21(1): 1–31. https://doi.org/10.1186/s13059-020-02154-5

Jing L, Mingcheng T, Feng Y (2005) Taxonomic reexamination of the genus *Chirolophis* in China waters. Chinese Journal of Oceanology and Limnology 23(2): 199–203. https://doi.org/10.1007/BF02894239

Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, De Bakker PI (2008) SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics 24(24): 2938–2939. https://doi.org/10.1093/bioinformatics/btn564

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M, Kawashima S, Katayama T, Araki M, Hirakawa M (2006) From genomics to chemical genomics: New developments in KEGG. Nucleic Acids Research 34(90001): D354–D357. https://doi.org/10.1093/nar/gkj102

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Kasai Y, Jindo T (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447(7145): 714–719. https://doi.org/10.1038/nature05846

Katoh K, Standley DM (2013) MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30(4): 772–780. https://doi.org/10.1093/molbev/mst010

Katti MV, Ranjekar PK, Gupta VS (2001) Differential distribution of simple sequence repeats in eukaryotic genome sequences. Molecular Biology and Evolution 18(7): 1161–1167. https://doi.org/10.1093/oxfordjournals.molbev.a003903

Koepfli KP, Paten B, O'Brien SJ (2015) The Genome 10K Project: A way forward. Annual Review of Animal Biosciences 3(1): 57–111. https://doi.org/10.1146/annurev-animal-090414-014900

Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research 35(9): 3100–3108. https://doi.org/10.1093/nar/gkm160

Li F, Chen X, Lu G, Qu J, Bian L, Chang Q, Ge J, Liu C, Zhang S, Chen S (2019a) Sequence and phylogenetic analysis of the mitochondrial genome for the Wolf-eel, *Anarrhichthys ocellatus* (Anarhichadidae: Perciformes). Mitochondrial DNA. Part B, Resources 4(2): 2884–2885. https://doi.org/10.1080/23802359.2019.1660260

Li H, Yi T, Gao L, Ma P, Zhang T, Yang J, Gitzendanner M, Fritsch P, Cai J, Luo Y, Wang H, van der Bank M, Zhang S-D, Wang Q-F, Wang J, Zhang Z-R, Fu C-N, Yang J, Hollingsworth PM, Chase MW, Soltis DE, Soltis PS, Li D-Z (2019b) Origin of angiosperms and the puzzle of the Jurassic gap. Nature Plants 5(5): 461–470. https://doi.org/10.1038/s41477-019-0421-0

Li Z, Tian C, Huang Y, Lin X, Wang Y, Jiang D, Zhu C, Chen H, Li G (2019c) A first insight into a draft genome of silver sillago (*Sillago sihama*) via genome survey sequencing. Animals (Basel) 9(10): 756. https://doi.org/10.3390/ani9100756

Liu B, Shi Y, Yuan J, Hu X, Zhang H, Li N, Li Z, Chen Y, Mu D, Fan W (2013) Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. ArXiv Preprint ArXiv:1308: 2012. https://doi.org/10.48550/arXiv.1308.2012

Liu Y, Xu P, Xu J, Huang Y, Liu Y, Fang H, Hu Y, You X, Bian C, Sun M, Gu R, Cui L, Zhang X, Xu P, Shi Q (2017) China is initiating the Aquatic 10-100-1,000 Genomics Program. Science China. Life Sciences 60(3): 329–332. https://doi.org/10.1007/s11427-016-0273-9

Liu F, Ma Z, Hou L, Diao Y, Wu W, Damm U, Song S, Cai L (2022) Updating species diversity of Colletotrichum, with a phylogenomic overview. Studies in Mycology, 1–86. https://doi.org/10.3114/sim.2022.101.01

Lohse M, Drechsel O, Kahlau S, Bock R (2013) Organellar Genome DRAW – A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Research 41(W1): W575–W581. https://doi.org/10.1093/nar/gkt289

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M (2005) Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Research 33(20): 6494–6506. https://doi.org/10.1093/nar/gki937

Longo GC, Lam L, Basnett B, Samhouri J, Hamilton S, Andrews K, Williams G, Goetz G, McClure M, Nichols KM (2020) Strong population differentiation in lingcod (*Ophiodon elongatus*) is driven by a small portion of the genome. Evolutionary Applications 13(10): 2536–2554. https://doi.org/10.1111/eva.13037

Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25(5): 955–964. https://doi.org/10.1093/nar/25.5.955

Margaryan A, Noer CL, Richter SR, Restrup ME, Bülow-Hansen JL, Leerhøi F, Langkjær EMR, Gopalakrishnan S, Carøe C, Gilbert MTP, Bohmann K (2021) Mitochondrial genomes of Danish vertebrate species generated for the national DNA reference database DNAmark. Environmental DNA 3(2): 472–480. https://doi.org/10.1002/edn3.138

Mu Y, Bian C, Liu R, Wang Y, Shao G, Li J, Qiu Y, He T, Li W, Ao J, Shi Q, Chen X (2021) Whole genome sequencing of a snailfish from the Yap Trench (~7,000 m) clarifies the molecular mechanisms underlying adaptation to the deep sea. PLOS Genetics 17(5): e1009530. https://doi.org/10.1371/journal.pgen.1009530

Nath S, Shaw DE, White MA (2021) Improved contiguity of the three-spine stickleback genome using long-read sequencing. G3 (Bethesda) 11: jkab007. https://doi.org/10.1093/g3journal/jkab007

Nie Y, Zhao H, Wang Z, Zhou Z, Liu X, Huang B (2021) The gene rearrangement loss transfer and deep intronic variation in mitochondrial genomes of *Conidiobolus*. Frontiers in Microbiology 12: 765733–765733. https://doi.org/10.3389/fmicb.2021.765733

Potter S, Consortium DTOL (2022) The genome sequence of the rock gunnel *Pholis gunnellus* (Linnaeus 1758). Wellcome Open Research 7: 58. https://doi.org/10.12688/wellcomeopenres.17699.1

Purcell CM, Seetharam AS, Snodgrass O, Ortega-García S, Hyde JR, Severin AJ (2018) Insights into teleost sex determination from the *Seriola dorsalis* genome assembly. BMC Genomics 19(1): 31. https://doi.org/10.1186/s12864-017-4403-1

Qiu BX, Fang SB, Ikhwanuddin M, Wong LL, Ma HY (2020) Genome survey and development of polymorphic microsatellite loci for *Sillago sihama* based on Illumina sequencing technology. Molecular Biology Reports 47(4): 3011–3017. https://doi.org/10.1007/s11033-020-05348-z

Ran JH, Shen TT, Wu H, Gong X, Wang X (2018) Phylogeny and evolutionary history of Pinaceae updated by transcriptomic analysis. Molecular Phylogenetics and Evolution 129: 106–116. https://doi.org/10.1016/j.ympev.2018.08.011

Sambrook J, Fritsch EF, Maniatis T (1982) Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press: New York NY, USA.

Sato M (1977) Histological observations on the cutaneous processes on the head of *Azuma emmnion* and *Hemitripterus villosus*. Japanese Journal of Ichthyology 24: 12–16. https://doi.org/10.11369/jji1950.24.12

Shiogaki M (1983) On the life history of the stichaeid fish *Chirolophis japonicus*. Japanese Journal of Ichthyology 29: 446–455. https://doi.org/10.11369/jji1950.29.446

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31(19): 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A, James TY, O'Donnell K, Roberson RW, Taylor TN, Uehling J, Vilgalys R, White MM, Stajich JE (2016) A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. Mycologia 108(5): 1028–1046. https://doi.org/10.3852/16-042

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30(9): 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19: ii215–ii225. https://doi.org/10.1093/bioinformatics/btg1080

Sun C, Li J, Dong J, Niu Y, Hu J, Lian J, Li W, Li J, Tian Y, Shi Q, Ye X (2021) Chromosome-level genome assembly for the largemouth bass *Micropterus salmoides* provides insights into adaptation to fresh and brackish water. Molecular Ecology Resources 21(1): 301–315. https://doi.org/10.1111/1755-0998.13256

Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L). Theoretical and Applied Genetics 106(3): 411–422. https://doi.org/10.1007/s00122-002-1031-0

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq-versatile and accurate annotation of organelle genomes. Nucleic Acids Research 45(W1): W6–W11. https://doi.org/10.1093/nar/gkx391

Wang K, Wang J, Zhu C, Yang L, Ren Y, Ruan J, Fan G, Hu J, Xu W, Bi X (2021) African lungfish genome sheds light on the vertebrate water-to-land transition. Cell 184(5): 1362–1376[e1318]. https://doi.org/10.1016/j.cell.2021.01.047

Wu CS, Ma ZY, Zheng GD, Zou SM, Zhang XJ, Zhang YA (2022) Chromosome-level genome assembly of grass carp (*Ctenopharyngodon idella*) provides insights into its genome evolution. BMC Genomics 23(1): 271. https://doi.org/10.1186/s12864-022-08503-x

Xu P, Zhang X, Wang X, Li J, Liu G, Kuang Y, Xu J, Zheng X, Ren L, Wang G, Zhang Y, Huo L, Zhao Z, Cao D, Lu C, Li C, Zhou Y, Liu Z, Fan Z, Shan G, Li X, Wu S, Song L, Hou G, Jiang Y, Jeney Z, Yu D, Wang L, Shao C, Song L, Sun J, Ji P, Wang J, Li Q, Xu L, Sun F, Feng J, Wang C, Wang S, Wang B, Li Y, Zhu Y, Xue W, Zhao L, Wang J, Gu Y, Lv W, Wu K, Xiao J, Wu J, Zhang Z, Yu J, Sun X (2014) Genome sequence and genetic diversity of the common carp *Cyprinus carpio*. Nature Genetics 46(11): 1212–1219. https://doi.org/10.1038/ng.3098

Xu S, Song N, Xiao S, Gao T (2020) Whole genome survey analysis and microsatellite motif identification of *Sebastiscus marmoratus*. Bioscience Reports 40(2): BSR20192252. https://doi.org/10.1042/BSR20192252

Yang H, Bao X, Wang B, Liu W (2016) The complete mitochondrial genome of *Chirolophis japonicus* (Perciformes: Stichaeidae). Mitochondrial DNA. Part A, DNA Mapping, Sequencing, and Analysis 27(6): 4419–4420. https://doi.org/10.3109/19401736.2015.1089552

Yang T, Huang X, Ning Z, Gao T (2021) Genome-Wide Survey Reveals the microsatellite characteristics and phylogenetic relationships of *Harpadon nehereu*. Current Issues in Molecular Biology 43(3): 1282–1292. https://doi.org/10.3390/cimb43030091

Yang T, Ning Z, Liu Y, Zhang S, Gao T (2022) Genome-wide survey and genetic characteristics of *Ophichthus evermanni* based on Illumina sequencing platform. Bioscience Reports 42(5): BSR20220460. https://doi.org/10.1042/BSR20220460

Zhao R, Lu Z, Cai S, Gao T, Xu S (2021) Whole genome survey and genetic markers development of crocodile flathead *Cociella crocodilus*. Animal Genetics 52(6): 891–895. https://doi.org/10.1111/age.13136

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA (2013) The MaSuRCA genome assembler. Bioinformatics 29(21): 2669–2677. https://doi.org/10.1093/bioinformatics/btt476

Supplementary material 1

K-mer analyses (K = 71) of *Chirolophis japonicus*, X-axis and Y-axis represent the K-mer depth and frequency for the corresponding depth

Authors: Lu Liu, Qi Liu, Tianxiang Gao

Data type: image

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1129.91543.suppl1

Supplementary material 2

Maximum likelihood phylogenomic tree of *Chirolophis japonicus*

Authors: Lu Liu, Qi Liu, Tianxiang Gao

Data type: phylogenomic tree

Explanation note: Maximum likelihood phylogenomic tree of *Chirolophis japonicus* based on amino acids of 694 single-copy genes.

Copyright notice: This dataset is made available under the Open Database License (http://opendatacommons.org/licenses/odbl/1.0/). The Open Database License (ODbL) is a license agreement intended to allow users to freely share, modify, and use this Dataset while maintaining this same freedom for others, provided that the original source and author(s) are credited.

Link: https://doi.org/10.3897/zookeys.1129.91543.suppl2